

عنوان مقاله:

خوشه بندی سلولی جریان داده با تعدد ابعاد

محل انتشار:

سومین کنفرانس داده کاوی (سال: 1388)

تعداد صفحات اصل مقاله: 10

نویسندگان:

تکتم دهقانی - دانشجوی کارشناسی ارشد، گروه کامپیوتر، دانشگاه آزاد اسلامی واحد مشهد

محمود نقیب زاده - استاد، گروه کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد

خلاصه مقاله:

در این مقاله روشی مقیاس پذیر از نظر تعدد ابعاد و اندازه مجموعه داده برای خوشه بندی روی خط جریان های داده ارائه شده است که در مقایسه با روش های پیشین علاوه بر مقیاس پذیری نسبت به تغییرات در اطلاعات تطابق پذیرتر و در شناسایی و تعیین خوشه ها دقیق تر و سریع تر است. در این روش در ابتدا فضای چند بعدی داده ها به سلول هایی با اندازه یکسان تقسیم می شود. در هر سلول توزیع آماری داده های اخیر که در محدوده ی آن سلول قرار دارند، ذخیره می شوند و بر اساس این اطلاعات، بدون نیاز به ذخیره سازی داده ها، خوشه بندی انجام می شود. سلول های پرتراکم به سلول های کوچکتر شکسته می شوند و این روند تا رسیدن به سلول پایه ادامه می یابد و سلول های خلوت برای کاهش حافظه ادغام می شوند. در این روش ساختاری کارا برای مدیریت سلول های در تمام ابعاد ارائه شده است، که دسترسی تصادفی و سریع به سلول ها را ممکن می سازد. ابتدا خوشه بندی یک بعدی انجام می شود، سپس خوشه ها با توجه به ارتباط بین توزیع داده ها در ابعاد مختلف، در یک روند پایین به بالا، با یکدیگر ترکیب و خوشه های نهایی تولید می شوند. با ذخیره سازی مرزهای دقیق خوشه ها در ابعاد مختلف، خوشه بندی دقیق تر انجام می شود و با اصلاح تعریف همسایگی زمان جستجو برای شناسایی همسایه های یک سلول که یکی از مشکلات اصلی خوشه بندی چندبعدی است، نیز کاهش می یابد. به منظور کاهش اثر داده های قدیمی در خوشه بندی، به اطلاعات وزنی اختصاص داده شده است و با گذشت زمان وزن آنها کاهش می یابد. در این روش خوشه بندی روی خط و تغییرات جریان در خوشه ها در نظر گرفته شده است.

کلمات کلیدی:

خوشه بندی، جریان داده، شبکه ی سلولی، مقیاس پذیری، خوشه بندی با ابعاد بالا

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/108981>

